

SMMRNA: a database of small molecule modulators of RNA

Ankita Mehta, Surabhi Sonam, Isha Gouri, Saurabh Loharch, Deepak K. Sharma and Raman Parkesh*

Department of Advanced Protein Science, Institute of Microbial Technology, Chandigarh-160036, India

Received August 15, 2013; Revised September 13, 2013; Accepted October 1, 2013

ABSTRACT

We have developed SMMRNA, an interactive database, available at <http://www.smmrna.org>, with special focus on small molecule ligands targeting RNA. Currently, SMMRNA consists of ~770 unique ligands along with structural images of RNA molecules. Each ligand in the SMMRNA contains information such as K_d , K_i , IC_{50} , ΔT_m , molecular weight (MW), hydrogen donor and acceptor count, XlogP, number of rotatable bonds, number of aromatic rings and 2D and 3D structures. These parameters can be explored using text search, advanced search, substructure and similarity-based analysis tools that are embedded in SMMRNA. A structure editor is provided for 3D visualization of ligands. Advance analysis can be performed using substructure and OpenBabel-based chemical similarity fingerprints. Upload facility for both RNA and ligands is also provided. The physicochemical properties of the ligands were further examined using OpenBabel descriptors, hierarchical clustering, binning partition and multidimensional scaling. We have also generated a 3D conformation database of ligands to support the structure and ligand-based screening. SMMRNA provides comprehensive resource for further design, development and refinement of small molecule modulators for selective targeting of RNA molecules.

INTRODUCTION

Recently, RNAs have been unraveled as unique molecules playing critical roles in developmental and physiological processes in all living organisms (1–5). RNA is involved in the progression of diseases such as infectious diseases (6–8) (e.g. HIV, AIDS, hepatitis C), metabolic diseases (9,10) (e.g. diabetes, cancer) and triplet repeat disorders

(e.g. myotonic dystrophy, Huntington's disease) (11–21). Druggability of RNAs has been validated in high-profile targets, for instance, ribosomal RNA (rRNA) could be targeted using aminoglycosides, macrolide, tetracycline and oxazolidinone (22–26). There are clinically approved antibiotics (e.g. erythromycin), which bind to RNA molecules. Structural analysis of antibiotics bound to ribosomal subunits has revealed that rRNA-small molecule recognition is mainly governed by electrostatic interaction, shape and hydrogen-bonding interactions (27–28). Recent success in crystallization of rRNA offers key structural information for RNA-based drug design (28–40). Diverse roles of RNAs provide numerous opportunities for specific targeting and modulating RNAs with small molecules.

Even though RNA adopts a favorable 3D structure, targeting of cellular RNA such as messenger RNA and microRNA for drug discovery was considered as complicated due to difficulty in crystallization, their conformational sampling and lack of RNA-specific computational tools. Excitingly, recent advances in biomedical and computational fields have provided vital information on RNA–small molecule interaction to overcome these issues to develop therapeutics for many diseases (41–44). Moreover, new developments in nuclear magnetic resonance and mass spectroscopy have made it possible to screen millions of small molecule compounds for finding selective inhibitors for specific RNA target (45–56).

Recently, number of RNA-based molecular targets have begun to grow rapidly with detailed elucidation of their structural and functional relationship (57–64). Also, small molecule inhibitors have been successfully developed for various different RNA molecules (65–75). Numerous publicly available databases are available providing information about RNA sequence, secondary structure and 3D RNA structure (76–86). To the best of our knowledge, there is no database that focuses on small molecule modulators along with their target RNA and experimentally determined binding data (K_d , K_i , IC_{50} , ΔT_m) of the corresponding RNA-inhibitor complex. The experimental

*To whom correspondence should be addressed. Tel: +91 172 6665488; Fax: +91 172 2690585; Email: rparkesh@imtech.res.in

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

data published in the literature are unstructured and difficult to navigate and as a result challenging to perform structure activity prediction. Moreover, chemical structures are depicted as ChemDraw images and are thus not searchable. The process of discovery of novel ligands to target RNA would be greatly facilitated by the availability of a small molecule database that is RNA-specific. In this regard, it would be highly valuable to collocate, organize and integrate RNA and their modulators along with experimental data in a publicly available domain that can be effectively navigated. SMMRNA is the first step in this direction. SMMRNA would facilitate the investigation of structure activity relationship between RNA and ligands, clustering of small molecules targeting RNA, provide information for fragment-based drug design and also aid in the quantitative structure activity relationship modeling of RNA.

MATERIALS AND METHODS

Data collection

The structural and experimental data were manually collected from peer-reviewed journals such as Biochemistry, Journal of American Chemical Society, Nature, Science and Journal of Molecular Biology. The relevant articles were selected from Pubmed, Web of Science and Google Scholar using keyword searches such as ‘small molecule targeting of RNA’, ‘small molecule RNA inhibitor’ and ‘targeting RNA with small molecules’. The details of RNA secondary structure, ligand structure, binding constant, IC₅₀, binding mode and assay performed were gathered from each article. The references provided within the articles were consulted to get further information about RNA secondary structures, ligand structures, experimental binding affinities and binding mode of various ligands. RNA structures were drawn as images using Adobe Photoshop (v4). Structures of ligands were drawn using ChemBioDraw Ultra (v11.0) and saved as sdf file format. In many cases, the chemical structure of ligand molecules, drawn as scaffold and abbreviated with R-group substituent in the published articles, was redrawn in computer-readable form. Before importing these structures into the database, structures were manually checked for atom valence and correctness of representation.

Database description

The database is built using Apache HTTP (web server) along with MySQL (database server). MySQL RDBMS (relational database management system) is used for storing data (see the Supplementary File S1 for a complete description of MySQL database). The general layout of SMMRNA showing homepage and search, browse, upload, literature and help tools are shown in Figure 1.

Substructure search

The command line version of VLifeMDS [VLifeMDS: Molecular Design Suite, VLife Sciences Technologies

Pvt. Ltd., Pune, India, 2013 (www.vlifesciences.com)] was used for performing a substructure search. The main steps involved in substructure search (as implemented in VLifeMDS) were: (i) Inputs were user-defined, consisting of template molecules to be matched against the reference molecule. Reference molecules were already stored in the database. (ii) Details of the molecules were built that include number of atoms (of different atom type; H, O, N, Cl, F and C), number of bonds (based on their types; single, double or triple), number of rings (3-, 4-, 5- and 6-member rings) and number of aromatic rings. (iii) The first step eliminates the substructure search based on the configuration set above. If any one of the configurations of template molecule was greater than a reference molecule, then no further processing was done. For instance, if the number of F atoms in the template was 3 and the number of F atoms in reference was 2, then obviously the template substructure did not match with that of the reference. (iv) An atom from template molecules was selected (this could be the atom with maximum number of connections). Next, the graph of reference molecule was traversed to find atom with the same degree and type as the template atom. (v) Each atom of the template was traversed recursively in a similar manner. If at any stage, no matching template atom was found in the reference molecule, the search was terminated. If all the atoms of the template molecule were matched with the reference molecule, the search was deemed successful.

2D structure editor for structure search

VLife2Draw (87) was used as an editor and viewer for searching 2D molecular structures in the SMMRNA database (Figure 3A). This is a Java applet and can be run in a browser with a suitable Java plugin. VLife2Draw allows the user to draw chemical structures, and to import and export these structures in molecule file formats (mol and sdf are supported formats). VLife2Draw provides basic chemical structure editing facility to draw bonds like single-, double-, triple-, stereo-up and stereo-down bonds. Templates for drawing 3–8-membered ring structures were also provided.

VLife2Draw applet runs on the client side, where a user can draw a structure that needs to be searched. After submitting the structure, applet converts these data into an sdf file format string, which is then transferred to the server. This information is captured by the server and stored in a temporary file. A Python script, written using VLifeMDS scripting framework, then uses this template molecule to search the SMMRNA database for ligand molecules that satisfy the structural similarity as indicated in the substructure search methodology.

Descriptor calculation

The command line version of VLifeMDS was used for calculating descriptors for ligands (Figure 2C). The descriptors computed by VLifeMDS are shown in Table 1.

3D structure viewing

The SMMRNA database contains ligand entries as 2D coordinates. For 3D depictions, the 2D coordinates are

Institute of Microbial Technology (सूक्ष्मजीव प्रौद्योगिकी संस्थान)
A Council of Scientific & Industrial Research (वैज्ञानिक औद्योगिक अनुसंधान परिषद)

About SMMRNA

RNA plays an important role in many regulatory processes such as transcription, protein synthesis and catalysis. RNA has complex three dimensional molecular structures like proteins, which can be exploited by small molecule ligands. Relatively moderate success has been achieved by targeting ribosomal RNA (rRNA).

[Continue Reading »](#)**About RNA**

RNA is a quite versatile macromolecule and plays an important role in various cellular processes. RNA is involved in important regulatory processes such as transcription, protein function and various catalytic activities. RNA show chemical similarity to DNA, however RNA molecules can fold to form complex three dimensional...

[Continue Reading »](#)**About Ligand**

The design and synthesis of novel, small therapeutic ligands has been an exciting area of pharmacology. The potentiality of macrolides (Erythromycin, Azithromycin), alkaloids (berberine, palmatine), aminoglycosides (Paromomycin, neomycin), tetracyclines (Doxycycline, Oxytetracycline), and oxazolidinones...

[Continue Reading »](#)

Figure 1. Homepage of SMMRNA database showing various analysis functionalities such as search, browse, upload, etc.

first curated for errors followed by the addition of appropriate number of hydrogen atoms. The structures are subsequently optimized using Merck molecular force field (MMFF), using a loose gradient cutoff (1.0 kcal/mol/angstroms). The optimized structures were further processed and stored into the database in a separate field in the MOL2 file format. Jmol molecular viewer was used for visualization of 3D structures of ligands (Jmol: an open-source Java viewer for chemical structures in 3D. <http://www.jmol.org/>). A standard ball-and-stick model is used to display the ligands (Figure 4A). Additionally, other structural displays such as Corey-Pauling-Koltun (CPK) spacefill, sticks and wireframe representation are possible by right clicking (using mouse or touchpad) and then selecting the renderMenu from the Jmol dialog box (Figure 4B). It is also possible to show various molecular surfaces such as Van der Waals, solvent accessible, dots, charge, etc (Figure 4C). CPK coloring scheme of atoms is used to depict the atom types; with the prominent atom types listed as Carbon—dark gray, Hydrogen—light gray, Oxygen—red and Nitrogen—green (92,93).

Clustering analysis

Hierarchical, binning and multidimensional cluster analysis of the database compounds were performed using the Chemmine software (<http://chemmine.ucr.edu/>) (94).

3D conformer generation

High-quality 3D conformers were generated using OMEGAv2.5 from openeye (<http://www.eyesopen.com/>) (95). Omega uses model building and torsional search for efficient conformational sampling. In the first step of model building, the chemical structure is fragmented along sigma bonds using the makefraglib utility of OMEGAv2.5, and then the structure is reconstructed by assembly of these fragments. These fragments were refined using modified MMFF94. In the second step, an ensemble of conformers was generated using torsional angle rules as described. For each molecule, 100 conformations were generated with energy cutoff of 10 kcal/mol. The conformational searches were terminated when the energy cutoff exceeded 10 kcal/mol or when the 100 structures were built (configuration parameter file is described in the Supplementary File S2).

RESULTS AND DISCUSSION**Organization and data retrieval of SMMRNA**

SMMRNA database is accessible via a user-friendly graphical user interface (GUI) at the web address www.smmrna.org. The interface allows textual search for both RNA and ligand (Figure 2), substructure and fingerprint search for ligands (Figure 3). In addition, browse facility is also provided to quickly locate ligands and RNA. The

A

Yeast t-RNA Phe

• RNA • Ligand

B

Search result for Yeast t-RNA Phe

RNA Id	RNA Name	RNA Image	Ligands	Expand	Collapse																																																																												
RNA0	Yeast t-RNA Phe		Click to Expand [-]																																																																														
<table border="1"> <thead> <tr> <th>Ligand Name</th> <th>K_i (μM)</th> <th>K_d (μM)</th> <th>IC₅₀ (μM)</th> </tr> </thead> <tbody> <tr> <td>Spermine</td> <td>-</td> <td>-</td> <td>8.5</td> </tr> <tr> <td>Ethidium bromide</td> <td>6X10⁶</td> <td>-</td> <td>-</td> </tr> <tr> <td>Methidiumpropyl-EDTA(MPE)</td> <td>1X10⁵-1X10⁶</td> <td>-</td> <td>-</td> </tr> <tr> <td>Hygromycin_B</td> <td>267</td> <td>-</td> <td>-</td> </tr> <tr> <td>Neomycin</td> <td>342</td> <td>-</td> <td>-</td> </tr> <tr> <td>Paromomycin_mol_mol</td> <td>2000</td> <td>-</td> <td>-</td> </tr> <tr> <td>Tobramycin</td> <td>8000</td> <td>-</td> <td>-</td> </tr> <tr> <td>Kanamycin_A</td> <td>5000</td> <td>-</td> <td>-</td> </tr> <tr> <td>KANAMYCIN_B</td> <td>11200</td> <td>-</td> <td>-</td> </tr> <tr> <td>ribostamycin</td> <td>23300</td> <td>-</td> <td>-</td> </tr> <tr> <td>sisomicin</td> <td>440</td> <td>-</td> <td>-</td> </tr> <tr> <td>5_episomicin</td> <td>2.5</td> <td>-</td> <td>-</td> </tr> <tr> <td>Lividomycin</td> <td>18000</td> <td>-</td> <td>-</td> </tr> <tr> <td>bleomycin A2</td> <td>-</td> <td>-</td> <td>-</td> </tr> <tr> <td>tris(1,10-phenanthroline)ruthenium(II)</td> <td>-</td> <td>-</td> <td>-</td> </tr> <tr> <td>TRIS(4,7,8-</td> <td>-</td> <td>-</td> <td>-</td> </tr> <tr> <td>TetramethylphenanthrolineRuthenium(II)</td> <td>-</td> <td>-</td> <td>-</td> </tr> <tr> <td>tris(4,7-diphenyl-1,10-phenanthroline)rhodium(II)</td> <td>-</td> <td>-</td> <td>-</td> </tr> </tbody> </table>						Ligand Name	K _i (μM)	K _d (μM)	IC ₅₀ (μM)	Spermine	-	-	8.5	Ethidium bromide	6X10 ⁶	-	-	Methidiumpropyl-EDTA(MPE)	1X10 ⁵ -1X10 ⁶	-	-	Hygromycin_B	267	-	-	Neomycin	342	-	-	Paromomycin_mol_mol	2000	-	-	Tobramycin	8000	-	-	Kanamycin_A	5000	-	-	KANAMYCIN_B	11200	-	-	ribostamycin	23300	-	-	sisomicin	440	-	-	5_episomicin	2.5	-	-	Lividomycin	18000	-	-	bleomycin A2	-	-	-	tris(1,10-phenanthroline)ruthenium(II)	-	-	-	TRIS(4,7,8-	-	-	-	TetramethylphenanthrolineRuthenium(II)	-	-	-	tris(4,7-diphenyl-1,10-phenanthroline)rhodium(II)	-	-	-
Ligand Name	K _i (μM)	K _d (μM)	IC ₅₀ (μM)																																																																														
Spermine	-	-	8.5																																																																														
Ethidium bromide	6X10 ⁶	-	-																																																																														
Methidiumpropyl-EDTA(MPE)	1X10 ⁵ -1X10 ⁶	-	-																																																																														
Hygromycin_B	267	-	-																																																																														
Neomycin	342	-	-																																																																														
Paromomycin_mol_mol	2000	-	-																																																																														
Tobramycin	8000	-	-																																																																														
Kanamycin_A	5000	-	-																																																																														
KANAMYCIN_B	11200	-	-																																																																														
ribostamycin	23300	-	-																																																																														
sisomicin	440	-	-																																																																														
5_episomicin	2.5	-	-																																																																														
Lividomycin	18000	-	-																																																																														
bleomycin A2	-	-	-																																																																														
tris(1,10-phenanthroline)ruthenium(II)	-	-	-																																																																														
TRIS(4,7,8-	-	-	-																																																																														
TetramethylphenanthrolineRuthenium(II)	-	-	-																																																																														
tris(4,7-diphenyl-1,10-phenanthroline)rhodium(II)	-	-	-																																																																														

C

Select	Property	Operator	Value
<input checked="" type="checkbox"/>	Molecular Weight	less than	500
<input type="checkbox"/>	Acceptor Count	select operator	Enter Value
<input type="checkbox"/>	Donor Count	select operator	Enter Value
<input type="checkbox"/>	XLOGP	select operator	Enter Value
<input type="checkbox"/>	Rotatable Bonds	select operator	Enter Value
<input type="checkbox"/>	Aromatic Rings	select operator	Enter Value
<input type="checkbox"/>	Kd Value	select operator	Enter Value
<input type="checkbox"/>	IC50	select operator	Enter Value

D

Search results for ligand having Molecular Weight < 500

Ligand Id	Ligand Name	Ligand Image	RNA	Expand	Collapse	[–]																																																
LIGAND103	Geneticin		RNA Name	K _i (μM)	K _d (μM)	IC ₅₀ (μM)																																																
<table border="1"> <thead> <tr> <th>RNA Name</th> <th>K_i (μM)</th> <th>K_d (μM)</th> <th>IC₅₀ (μM)</th> </tr> </thead> <tbody> <tr> <td>T4 phage derived Id intron</td> <td>-</td> <td>-</td> <td>1000</td> </tr> <tr> <td>Eubacterial A-site_sec str of A-site from E.coli 16S(rRNA)</td> <td>-</td> <td>5X10⁴</td> <td>-</td> </tr> <tr> <td>Bcr-Ab1-mRNA</td> <td>-</td> <td>20</td> <td>-</td> </tr> <tr> <td>PAX3-FKHR-mRNA</td> <td>-</td> <td>30</td> <td>-</td> </tr> </tbody> </table>							RNA Name	K _i (μM)	K _d (μM)	IC ₅₀ (μM)	T4 phage derived Id intron	-	-	1000	Eubacterial A-site_sec str of A-site from E.coli 16S(rRNA)	-	5X10 ⁴	-	Bcr-Ab1-mRNA	-	20	-	PAX3-FKHR-mRNA	-	30	-																												
RNA Name	K _i (μM)	K _d (μM)	IC ₅₀ (μM)																																																			
T4 phage derived Id intron	-	-	1000																																																			
Eubacterial A-site_sec str of A-site from E.coli 16S(rRNA)	-	5X10 ⁴	-																																																			
Bcr-Ab1-mRNA	-	20	-																																																			
PAX3-FKHR-mRNA	-	30	-																																																			
LIGAND104	Tobramycin		RNA Name	K _i (μM)	K _d (μM)	IC ₅₀ (μM)																																																
<table border="1"> <thead> <tr> <th>RNA Name</th> <th>K_i (μM)</th> <th>K_d (μM)</th> <th>IC₅₀ (μM)</th> </tr> </thead> <tbody> <tr> <td>Yeast t-RNA Phe</td> <td>8000</td> <td>-</td> <td>-</td> </tr> <tr> <td>T4 phage derived Id intron decoding region 16S(rRNA</td> <td>-</td> <td>1.689-0.16</td> <td>-</td> </tr> <tr> <td>Bacterial ribosomal A-site</td> <td>-</td> <td>0.08-0.12</td> <td>-</td> </tr> <tr> <td>16S rRNA A SITE</td> <td>-</td> <td>1-2</td> <td>-</td> </tr> <tr> <td>T box antiterminator RNA</td> <td>-</td> <td>38</td> <td>-</td> </tr> <tr> <td>SLI of Rep A</td> <td>-</td> <td>0.04</td> <td>-</td> </tr> <tr> <td>Thymidilate synthase-mRNA</td> <td>-</td> <td>2.2</td> <td>-</td> </tr> <tr> <td>Bcr-Ab1-mRNA</td> <td>-</td> <td>3.8</td> <td>-</td> </tr> <tr> <td>PAX3-FKHR-mRNA</td> <td>-</td> <td>1.8</td> <td>-</td> </tr> <tr> <td>RRE-RNA</td> <td>-</td> <td>3</td> <td>45</td> </tr> <tr> <td>RRE RNA</td> <td>-</td> <td>-</td> <td>3</td> </tr> </tbody> </table>							RNA Name	K _i (μM)	K _d (μM)	IC ₅₀ (μM)	Yeast t-RNA Phe	8000	-	-	T4 phage derived Id intron decoding region 16S(rRNA	-	1.689-0.16	-	Bacterial ribosomal A-site	-	0.08-0.12	-	16S rRNA A SITE	-	1-2	-	T box antiterminator RNA	-	38	-	SLI of Rep A	-	0.04	-	Thymidilate synthase-mRNA	-	2.2	-	Bcr-Ab1-mRNA	-	3.8	-	PAX3-FKHR-mRNA	-	1.8	-	RRE-RNA	-	3	45	RRE RNA	-	-	3
RNA Name	K _i (μM)	K _d (μM)	IC ₅₀ (μM)																																																			
Yeast t-RNA Phe	8000	-	-																																																			
T4 phage derived Id intron decoding region 16S(rRNA	-	1.689-0.16	-																																																			
Bacterial ribosomal A-site	-	0.08-0.12	-																																																			
16S rRNA A SITE	-	1-2	-																																																			
T box antiterminator RNA	-	38	-																																																			
SLI of Rep A	-	0.04	-																																																			
Thymidilate synthase-mRNA	-	2.2	-																																																			
Bcr-Ab1-mRNA	-	3.8	-																																																			
PAX3-FKHR-mRNA	-	1.8	-																																																			
RRE-RNA	-	3	45																																																			
RRE RNA	-	-	3																																																			

Figure 2. Examples of text and advanced property-based search in SMMRNA. (A) Screenshot showing dialog box for RNA and ligand text-based search. (B) Screenshot showing results based on RNA text search, e.g. using yeast t-RNA as text string, the information about the RNA and various ligands that target this particular RNA is depicted. (C) Screenshot showing dialog box for ligand search based on advanced properties such as MW, IC₅₀, hydrogen bond donor, acceptor count, etc. (D) Screenshot showing results based on MW cutoff of <500.

database also allows ligand-based substructure and fingerprint search (Figure 3). The ligands can be searched by their common or IUPAC names. For a molecule, this search would retrieve all the RNAs where the molecule acts as modulator together with related experimental data. Additionally for each compound, 2D and 3D structures, along with other calculated properties, appear in the drop down menu. Both 2D and 3D structures of the ligand can be downloaded (Figure 4D). For 3D representations of ligands, 2D structure was first corrected for errors, and then an appropriate number of hydrogen atoms were added. These 3D structures are energy minimized with MMFFs using gradient convergence of 1.0 kcal/(molÅ). The energy-optimized structures are further processed and stored into the database in a separate field in the MOL2 file format.

Online tools for structure and similarity search

We have implemented many web-based tools for structure-based search and the similarity analysis of various molecules reported in the SMMRNA. These tools are as described later in the text.

Advanced search

Under the search menu, an advanced tab is inserted to explore ligands using physicochemical properties such as molecular weight, XlogP and number of rotatable bonds (Figure 2C). The data generated can be filtered using a specific property or a combination of properties such as acceptor count, XlogP, etc. Each property could be controlled by an operator (greater than, less than or range) and a subsequent value for filtering. Users can further limit their search by applying a filter based on RNA names, if they are interested only in a single RNA or a class of RNA molecules.

Substructure search

A structure-based search could be easily performed by using powerful editors embedded within the web page. The user can draw a complete or partial structure of the query ligand to retrieve similar compounds (Figure 3). VLife2Draw applet runs on the client side, where a user can draw a structure that needs to be searched. After submitting the structure, applet converts these data into an sdf format string, which is then transferred to the

Table 1. Various descriptors along with their description and significance used for advanced search criteria

Descriptor	Comments
Acceptor count	Number of hydrogen bond acceptors in the molecule (88).
Donor count	Number of hydrogen bond donors in the molecule (88).
Rotatable bonds	Provides information about conformation sampling of a molecule, which is useful in docking. This is important in docking, where a particular conformation might show favorable binding.
Aromatic rings	Number of aromatic rings in a molecule. Generally, the presence of aromatic rings in a molecule result in the possibility of $\pi-\pi$ interactions prevalent in binding sites.
XlogP	Distribution-coefficient of the inhibitors, drug like properties and lipophilic efficiency (89–91).

A Substructure Search Similarity Search

Upload Template(.mol2 or .sdf): Choose File No file chosen Search template

Enter SMILES: Enter SMILES Search SMILES

Draw structure:

Submit Structure

B Substructure Search Similarity Search

Upload Template(.mol2 or .sdf): Choose File No file chosen Search template

Enter SMILES: Enter SMILES Search SMILES

Draw structure:

Submit Structure

C

Ligand Id	Ligand Name	Ligand Image	RNA	Expand Collapse
LIGAND280	1,4-bis((1H-indol-3-yl)methyl)piperazine		Click to Expand [-]	
			RNA Name K _i (μM) K _d (μM) IC ₅₀ (μM)	
			A-site 16S rRNA_E_coli (1) - - 216 -	
			3D View [+] Information [+] Downloads [+]	
LIGAND149	2,6-diaminopurine		Click to Expand [-]	
			RNA Name K _i (μM) K _d (μM) IC ₅₀ (μM)	
			ADENINE RIBOSWITCH - 3 -	
			3D View [+] Information [+] Downloads [+]	

D Top 10 ligands based on similarity (Tanimoto coefficient)

Ligand Id	Ligand Name	Ligand Image	Tanimoto	RNA	Expand Collapse
LIGAND184	Malachite green		0.167	Click to Expand [-]	
			RNA Name K _i (μM) K _d (μM) IC ₅₀ (μM)		
			Aptamer 0.05 -		
LIGAND748	SL3 RNA_ligand4		0.156	Click to Expand [-]	
			RNA Name K _i (μM) K _d (μM) IC ₅₀ (μM)		
			SL3 RNA 1.1 ± 0.1 -		
			Double stranded RNA seq 3.2 ± 0.5 -		
			single stranded RNA seq 2.2 -		
			SL3 RNA_Ncp7 Complex 200 ± 100 - 700 ± 350		

Figure 3. Examples of substructure and chemical similarity-based search in SMMRNA. Screenshots of (A) substructure query, (B) similarity search and (C and D) the data obtained using substructure and similarity-based search, respectively.

server. The server side script captures this information and stores it into a temporary file. A Python script, written using VLifeMDS scripting framework, then uses this template molecule to search the SMMRNA database for ligand molecules that satisfy the structural similarity as indicated in the substructure search methodology. The steps implemented in VLifeMDS for performing a

substructure search are as follows: (i) user-defined input to be matched against the database molecule; (ii) molecule generation using atoms (H, O, N and C), bonds (single, double and triple) and carbocyclic, heterocyclic and aromatic rings of different sizes (3-, 4-, 5-, 6-membered); (iii) atom template matching between the input and the database reference molecules; and (iv) graph matching

A

Ligand Id	Name	Image
LIGAND37	acridine orange	

B

C

D

Ligand Id	Name	Image																
LIGAND37	acridine orange	 <table border="1"> <thead> <tr> <th>Property</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>Molecule Weight</td> <td>179.22</td> </tr> <tr> <td>Acceptor count</td> <td>0</td> </tr> <tr> <td>Donor count</td> <td>0</td> </tr> <tr> <td>XLOGP</td> <td>3.5</td> </tr> <tr> <td>Rotatable bond</td> <td>0</td> </tr> <tr> <td>Aromatic rings</td> <td>3</td> </tr> <tr> <td>References</td> <td>Lacourciere, K. A.et.al., Biochemistry 2000,39, 5630</td> </tr> </tbody> </table> Download 2D Structure Download 3D Structure	Property	Value	Molecule Weight	179.22	Acceptor count	0	Donor count	0	XLOGP	3.5	Rotatable bond	0	Aromatic rings	3	References	Lacourciere, K. A.et.al., Biochemistry 2000,39, 5630
Property	Value																	
Molecule Weight	179.22																	
Acceptor count	0																	
Donor count	0																	
XLOGP	3.5																	
Rotatable bond	0																	
Aromatic rings	3																	
References	Lacourciere, K. A.et.al., Biochemistry 2000,39, 5630																	

Figure 4. SMMRNA ligand screenshot showing (A) 2D and 3D structural view of the ligand, (B) various 3D structural representations using Jmol dialog box, (C) various molecular surface representations such as dots, Van der Waals, charge, etc and (D) advanced property, reference and comment information and 2D and 3D structure download links.

between input and reference molecule to establish connectivity. If there is a match between atom template and graph connectivity between input and reference molecule, the search would return with the matching list of structures. Graph matching is based on graph theory, which involves modeling the pairwise similarity between two molecules. In VLifeMDS, molecules are represented by graphs, which consist of vertices (atom) and edges (bond connecting atoms). Vertices are pictorially represented as ‘dots’ or ‘circles’, whereas the edges are represented by drawing an arc. Mathematically, a graph can be represented by a matrix of atoms and bonds, and each entry (atom, bonds) contains the bond endpoint data. For substructure search matching, all it involves is to determine which graphs in a database are subgraph isomorphic to a query graph (see the Supplementary File S3 for illustration of graph matching).

Similarity search

Ligand similarity search of SMMRNA is quite versatile, as it has the capability to identify as well as rank the structures based on the query. A fingerprint represents a predefined fragment or feature found in a molecule. The fingerprint search uses a binary fingerprint for the chemical structure. A fingerprint is a series of binary (1/0) bits that are arranged in sequence. Each bit position corresponds to certain information such as the presence or absence of an atom, type of ring, element count and a substructure. During a substructure similarity search, a fingerprint of the query is computed initially. Tanimoto coefficient is based on the fingerprints, which are binary vectors, with a value of 1 and 0, with 1 indicating the ‘presence of’ and 0 indicating the ‘absence of’ fingerprints. Tanimoto coefficient quantifies the similarity between query and database

molecules and has a similarity score range from 0.0 to 1.0. For computing Tanimoto coefficient, each ligand in the SMMRNA has a pre-computed fingerprint incorporated. The generated fingerprint is then matched with every entry in the database, and a similarity score, which lies between 0.0 and 1.0, is computed. The structures are sorted in descending order relative to similarity score, with the top hits representing the most similar structure to the search query (Figure 3D). SMMRNA uses OpenBabel fingerprints (<http://openbabel.org/>) search for similarity analysis (96).

Significance of the database

We have reviewed >900 original research articles and collected all the relevant information. In SMMRNA, we have not included the PubChem Bioassay data to avoid duplication and problems about the credibility and validation of high-throughput screening (HTS) data. The SMMRNA database contains unique molecules and includes only those molecules for which the experimental data are reported in the peer-reviewed literature. In few articles, experimental parameters were not reported, although the binding to RNA molecules was confirmed and validated with various spectroscopic and other binding techniques such as gel shift and radioactivity assay. We have included these molecules and annotated comments and references to highlight these observations so that the user can easily verify and cross validate.

The SMMRNA database is very diverse and includes molecular classes such as antibiotics, peptides, amino acids, modified amino acids, dyes, nucleotides, polycations, intercalators, inorganic and organometallic compounds, which are unique to this database (see the Supplementary File S4). One particular observation is that many antibiotics act as ligands for RNA. Literature analysis reveals (97) that there are eight important clinically approved antibiotics, whose mechanism of action involves RNA binding, and the research and development efforts on antibiotics in recent years have severely declined. The data provided here are appealing for antibiotics research and would aid in the design of improved or completely new antibiotics (using various computational methods).

Chemical space analysis of database molecules

RNA is a highly charged molecule and possesses a high degree of conformational flexibility. Thus, it is important in the early drug discovery stages to predict and optimize the favorable physicochemical properties of ligands to facilitate their efficient binding to RNA molecules. To understand whether various filters that are commonly applied to other biological targets are effective, we analyzed the molecules in the SMMRNA database using standard filters such as Lipinski Rule of Five, ZINC and PAINS filter (98–100). ZINC filter is less restricted compared with other filters and uses relaxed cutoff value of various physicochemical parameters such as molecular weight, hydrogen bond donor and acceptor count, logP and number of rotatable bonds (see the Supplementary File S5 for ZINC filters used). About two-third of the molecules failed to pass these filters. These results are

important in understanding the chemical space, which we should focus in developing ligands for targeting RNA. For targeting RNA, specific physicochemical parameters such as more hydrophilicity, lower lipophilicity (especially for developing new antibiotics), high molecular weight and larger polar surface area are required. Most of the ligands that bind to RNA are highly charged, e.g. polycation derivatives and aminoglycosides. However, there are small drug-like molecules such as DAPI, etc, which can be optimized to bind RNA with varying binding affinity. It is thus possible to develop lead-like molecules to target RNA by modification of these molecules using organic synthesis. From the analysis of the molecules in SMMRNA, it is clear that we need to develop separate filter parameters for analyzing ligands targeting RNA. In particular, we need to be flexible in terms of hydrogen bond donor, hydrogen bond acceptor, number of rotatable bonds, ring size to design and synthesize diverse and optimized ligands of higher selectivity and sensitivity. Also, majority of RNA targeting molecules have significantly higher topological polar surface area as compared with the drug-like molecules, where the value is around $<140\text{ \AA}^2$. These results are highly important, as recent reviews (97) have highlighted the need to develop entirely different physicochemical parameters for targeting RNA.

Clustering analysis

Clustering is an important tool for medicinal or computational chemist and is widely used in application ranging from drug discovery to lead optimization (101,102). Clustering techniques are used to identify molecule scaffolds that share common structural properties (102). It is widely accepted that structurally similar molecules will show similar biological activity profile (103). ChemMine tools for clustering analysis include hierarchical clustering, multidimensional scaling (MDS) and binning clustering. We have used these clustering tools to identify common scaffolds based on structural and molecular descriptors. This will lead to the identification of potential molecular entities for structure activity relationship. We have performed hierarchical, binning and MDS to identify and visualize the chemical space of ligand molecules.

We have performed agglomerative hierarchical clustering method, which involves a progressive combination of the most similar molecules into related groups, and the results are shown as a dendrogram, highlighting the relationship between clusters. It has the advantage that any number of clusters can be selected. However, it is difficult to visualize the chemical space covered by database molecules. For visualization of chemical space, we performed binning partitioning clustering and MDS. Bin-based partitioning methods are useful for the identification of underrepresented chemical space, which can be fulfilled by synthesizing more compounds. In bin-based partitioning method, we subdivided multidimensional chemical space into various bins using a Tanimoto similarity cutoff of 0.4. Compounds falling into the same bin volumes are deemed to have similar chemical properties. Bin-based partitioning of database compounds, which

passed the ZINC filter resulted in the creation of bins with variable bin sizes. We observed many bin clusters with sizable number of similar molecules. For example, there is an interesting cluster of antibiotics that target RNA molecules. We observed a number of bin clusters represented by just single molecular scaffold. We believe that greater medicinal chemistry effort should be made to expand the compounds representing these particular scaffolds.

3D conformation

For high-throughput virtual screening, such as ligand and structure-based drug design, we generated a multiconformer database of 647 compounds using MMFF94. For each compound, 100 conformers were generated specifying the energy cutoff of 10 kcal/mol and root-mean-square deviation of 0.5. For some compounds, the number of conformers generated was less than the specified as a result of energy and root-mean-square deviation criteria. Around 128 compounds failed as a result of the criteria specified. We believe this multiconformer database (with 33 866 conformations, see the Supplementary File S6) will be very useful, as it provides the structural information about the conformational states of all the compounds, which is vital for *in silico* drug designing.

CONCLUSIONS AND FUTURE PROSPECTS

We have developed SMMRNA (www.smmrna.org), a chemoinformatics platform to facilitate interactive exploring of RNA molecules and their modulators. The database comprises structural images of RNA, chemical structure of ligands, related experimental data (K_d , K_i , IC₅₀, ΔT_m values), 3D conformers and chemoinformatic information. The database can be browsed using RNA names, ligand structures, substructures and fingerprint-based chemical similarity search (see the Supplementary File S7, which shows the video tutorial for navigating SMMRNA database). SMMRNA would facilitate structure activity relationship studies, statistical analysis, fragment-based drug design, virtual screening and molecular docking studies to assist in developing RNA-based small molecule modulators (for comparison with other databases see the Supplementary File S8). In general, this publicly available database would be beneficial for the design and discovery of modulators targeting RNA-mediated diseases such as diabetes, neurodegenerative disorders and cardiovascular disorders. We are continually reviewing the older literature, patents, etc, so as to include additional experimentally validated ligand entries. We have also provided the upload facility for both ligand and RNA entries for other researchers so that SMMRNA will continue to grow and thus provide comprehensive RNA-ligand data. In future versions, we would include RNA structure drawing tool, RNA alignment tool, fragment generation capability, 3D structure comparison, automatic docking protocol and principal component analysis. The main aim of the database is to provide computational and medicinal chemistry tools for

research community who are interested in medicinal chemistry, biology and biochemistry of RNA. We believe that the set of tools provided in SMMRNA would invite scientists from diverse backgrounds to initiate drug design and development of small molecule modulators for targeting various RNA molecules.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr Rethi Madathil for help in preparing the manuscript, Ms Isha for drawing additional structures and clustering analysis and Dr Kundan, Dr Ganesh and Mr Kiran for technical help.

FUNDING

The Institute of Microbial Technology [grant Infra 62 to R.P.]; Institute of Microbial Technology Research Internship (to A.M. and S.S.) and CSIR JRF (to S.L.); Recipient of Ramalingaswami fellowship (to D.K.S.). Funding for open access charge: Intramural research funds of Institute of Microbial Technology, Chandigarh.

Conflict of interest statement. None declared.

REFERENCES

1. Moore,P.B. and Steitz,T.A. (2011) The roles of RNA in the synthesis of protein. *Cold Spring Harb. Perspect. Biol.*, **3**, a003780.
2. Walsh,C. (2000) Molecular mechanisms that confer antibacterial drug resistance. *Nature*, **406**, 775.
3. Hermann,T. and Westhof,E. (1998) RNA as a drug target: chemical, modelling, and evolutionary tools. *Curr. Opin. Biotechnol.*, **9**, 66–73.
4. Blanchard,S.C., Gonzalez,R.L., Kim,H.D., Chu,S. and Puglisi,J.D. (2004) tRNA selection and kinetic proofreading in translation. *Nat. Struct. Mol. Biol.*, **11**, 1008–1014.
5. Noller,H.F. (1991) Drugs and the RNA world. *Nature*, **353**, 302–303.
6. Zapp,M.L., Stern,S. and Green,M.R. (1993) Small molecules that selectively block RNA binding of HIV-1 rev protein inhibit rev function and viral production. *Cell*, **74**, 969–978.
7. Malim,M.H. and Cullen,B.R. (1991) HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell*, **65**, 241–248.
8. Jensen,K.B., Green,L., MacDougal-Waugh,S. and Tuerk,C. (1994) Characterization of an *in vitro*-selected RNA ligand to the HIV-1 rev protein. *J. Mol. Biol.*, **235**, 237–247.
9. Cooper,T.A., Wan,L. and Dreyfuss,G. (2009) RNA and disease. *Cell*, **136**, 777–793.
10. Pérez,B., Ugarte,M. and Desviat,L. (2012) In: Erdmann,V.A. and Barciszewski,J. (eds), *From Nucleic Acids Sequences to Molecular Medicine*. Springer Verlag, Berlin/Heidelberg, Germany, pp. 357–370.
11. Fiszer,A., Mykowska,A. and Krzyzosiak,W.J. (2011) Inhibition of mutant huntingtin expression by RNA duplex targeting expanded CAG repeats. *Nucleic Acids Res.*, **39**, 5578–5585.
12. Galka-Marciniak,P., Urbanek,M.O. and Krzyzosiak,W.J. (2012) Triplet repeats in transcripts: structural insights into RNA toxicity. *Biol. Chem.*, **393**, 1299–1315.
13. Hu,J., Liu,J., Yu,D., Chu,Y. and Corey,D.R. (2012) Mechanism of allele-selective inhibition of huntingtin expression by duplex

- RNAs that target CAG repeats: function through the RNAi pathway. *Nucleic Acids Res.*, **40**, 11270–11280.
14. Kiliszek,A., Kierzek,R., Krzyzosiak,W.J. and Rypniewski,W. (2011) Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Res.*, **39**, 7308–7315.
 15. Cooper,T.A. (2009) Chemical reversal of the RNA gain of function in myotonic dystrophy. *Proc. Natl Acad. Sci. USA*, **106**, 18433–18434.
 16. Warf,M.B., Nakamori,M., Matthys,C.M., Thornton,C.A. and Berglund,J.A. (2009) Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl Acad. Sci. USA*, **106**, 18551–18556.
 17. Wheeler,T.M., Leger,A.J., Pandey,S.K., MacLeod,A.R., Nakamori,M., Cheng,S.H., Wentworth,B.M., Bennett,C.F. and Thornton,C.A. (2012) Targeting nuclear RNA for *in vivo* correction of myotonic dystrophy. *Nature*, **488**, 111–115.
 18. Sicot,G., Gourdon,G. and Gomes-Pereira,M. (2011) Myotonic dystrophy, when simple repeats reveal complex pathogenic entities: new findings and future challenges. *Hum. Mol. Genet.*, **20**, 5.
 19. Parkesh,R., Childs-Diasney,J.L., Nakamori,M., Kumar,A., Wang,E., Wang,T., Hoskins,J., Tran,T., Housman,D., Thornton,C.A. *et al.* (2012) Design of a bioactive small molecule that targets the myotonic dystrophy type 1 RNA via an RNA motif-ligand database and chemical similarity searching. *J. Am. Chem. Soc.*, **134**, 4731–4742.
 20. Watson,L.M. and Wood,M.J. (2012) RNA therapy for polyglutamine neurodegenerative diseases. *Expert Rev. Mol. Med.*, **31**, 1.
 21. Nakamori,M., Gourdon,G. and Thornton,C.A. (2011) Stabilization of expanded (CTG)*(CAG) repeats by antisense oligonucleotides. *Mol. Ther.*, **19**, 2222–2227.
 22. Fischbach,M.A. and Walsh,C.T. (2009) Antibiotics for emerging pathogens. *Science*, **325**, 1089–1093.
 23. Brandi,L., Fabbretti,A., Milon,P., Carotti,M., Pon,C.L. and Gualerzi,C.O. (2007) Methods for identifying compounds that specifically target translation. *Methods Enzymol.*, **431**, 229–267.
 24. Lange,R.P., Locher,H.H., Wyss,P.C. and Then,R.L. (2007) The targets of currently used antibacterial agents: lessons for drug discovery. *Curr. Pharm. Des.*, **13**, 3140–3154.
 25. Kohanski,M.A., Dwyer,D.J. and Collins,J.J. (2010) How antibiotics kill bacteria: from targets to networks. *Nat. Rev. Microbiol.*, **8**, 423–435.
 26. McCoy,L.S., Xie,Y. and Tor,Y. (2011) Antibiotics that target protein synthesis. *Wiley Interdiscip. Rev. RNA*, **2**, 209–232.
 27. Harvey,I., Garneau,P. and Pelletier,J. (2002) Inhibition of translation by RNA–small molecule interactions. *RNA*, **8**, 452–463.
 28. Cate,J.H., Yusupov,M.M., Yusupova,G.Z., Earnest,T.N. and Noller,H.F. (1999) X-ray crystal structures of 70S ribosome functional complexes. *Science*, **285**, 2095–2104.
 29. Clemons,W.M. Jr, May,J.L., Wimberly,B.T., McCutcheon,J.P., Capel,M.S. and Ramakrishnan,V. (1999) Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature*, **400**, 833–840.
 30. Brodersen,D.E., Clemons,W.M. Jr, Carter,A.P., Morgan-Warren,R.J., Wimberly,B.T. and Ramakrishnan,V. (2000) The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell*, **103**, 1143–1154.
 31. Carter,A.P., Clemons,W.M., Brodersen,D.E., Morgan-Warren,R.J., Wimberly,B.T. and Ramakrishnan,V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, **407**, 340–348.
 32. Wimberly,B.T., Brodersen,D.E., Clemons,W.M. Jr, Morgan-Warren,R.J., Carter,A.P., Vonrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
 33. Clemons,W.M. Jr, Brodersen,D.E., Morgan-Warren,R.J., Hartsch,T., Wimberly,B.T. and Ramakrishnan,V. (2001) Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science*, **291**, 498–501.
 34. Yonath,A. (2002) High-resolution structures of large ribosomal subunits from mesophilic eubacteria and halophilic archaea at various functional States. *Curr. Protein. Pept. Sci.*, **3**, 67–78.
 35. Ogle,J.M., Brodersen,D.E., Clemons,W.M. Jr, Tarry,M.J., Carter,A.P. and Ramakrishnan,V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science*, **292**, 897–902.
 36. Ogle,J.M., Carter,A.P. and Ramakrishnan,V. (2003) Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem. Sci.*, **28**, 259–266.
 37. Schmeing,T.M. and Ramakrishnan,V. (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature*, **461**, 1234–1242.
 38. Moore,P.B. and Steitz,T.A. (2003) The structural basis of large ribosomal subunit function. *Annu. Rev. Biochem.*, **72**, 813–850.
 39. Steitz,T.A. and Moore,P.B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.*, **28**, 411–418.
 40. Steitz,T.A. (2008) A structural understanding of the dynamic ribosoma machine. *Nat. Rev. Mol. Cell Biol.*, **9**, 242–253.
 41. Dieterich,C. and Stadler,P.F. (2013) Computational biology of RNA interactions. *Wiley Interdiscip. Rev. RNA*, **4**, 107–120.
 42. Chou,C.H., Lin,F.M., Chou,M.T., Hsu,S.D., Chang,T.H., Weng,S.L., Shrestha,S., Hsiao,C.C., Hung,J.H. and Huang,H.D. (2013) A computational approach for identifying microRNA-target interactions using high-throughput CLIP and PAR-CLIP sequencing. *BMC Genomics*, **14**(Suppl.1), S2.
 43. Kuzu,G., Keskin,O., Gursoy,A. and Nussinov,R. (2012) Expanding the conformational selection paradigm in protein–ligand docking. *Methods Mol. Biol.*, **819**, 59–74.
 44. Sim,A.Y., Minary,P. and Levitt,M. (2012) Modeling nucleic acids. *Curr. Opin. Struct. Biol.*, **22**, 273–278.
 45. Moumne,R., Catala,M., Larue,V., Micouin,L. and Tisne,C. (2012) Fragment-based design of small RNA binders: promising developments and contribution of NMR. *Biochimie*, **94**, 1607–1619.
 46. Furtig,B., Richter,C., Wohner,J. and Schwalbe,H. (2003) NMR spectroscopy of RNA. *ChemBioChem*, **4**, 936–962.
 47. Qin,P.Z. and Dieckmann,T. (2004) Application of NMR and EPR methods to the study of RNA. *Curr. Opin. Struct. Biol.*, **14**, 350–359.
 48. Furtig,B., Buck,J., Manoharan,V., Bermel,W., Jaschke,A., Wenter,P., Pitsch,S. and Schwalbe,H. (2007) Time-resolved NMR studies of RNA folding. *Biopolymers*, **86**, 360–383.
 49. Huber,W. (2005) A new strategy for improved secondary screening and lead optimization using high-resolution SPR characterization of compound-target interactions. *J. Mol. Recognit.*, **18**, 273–281.
 50. Zhu,Y., Valdes,R. Jr and Jortani,S.A. (2005) Application of bioaffinity mass spectrometry for analysis of ligands. *Ther Drug Monit.*, **27**, 694–699.
 51. Schug,K.A. (2007) Solution phase enantioselective recognition and discrimination by electrospray ionization—mass spectrometry: state-of-the-art, methods, and an eye towards increased throughput measurements. *Comb. Chem. High Throughput Screen*, **10**, 301–316.
 52. Oeffinger,M. (2012) Two steps forward—one step back: advances in affinity purification mass spectrometry of macromolecular complexes. *Proteomics*, **12**, 1591–1608.
 53. Tuzimski,T. (2011) Application of different modes of thin-layer chromatography and mass spectrometry for the separation and detection of large and small biomolecules. *J. Chromatogr. A*, **1218**, 8799–8812.
 54. Nakayama,H., Takahashi,N. and Isobe,T. (2011) Informatics for mass spectrometry-based RNA analysis. *Mass Spectrom. Rev.*, **30**, 1000–1012.
 55. Krug,K., Nahnsen,S. and Macek,B. (2011) Mass spectrometry at the interface of proteomics and genomics. *Mol. Biosyst.*, **7**, 284–291.
 56. Giessing,A.M. and Kirpekar,F. (2012) Mass spectrometry in the biology of RNA and its modifications. *J. Proteomics*, **75**, 3434–3449.
 57. DeJong,E.S., Luy,B. and Marino,J.P. (2002) RNA and RNA–protein complexes as targets for therapeutic intervention. *Curr. Top. Med. Chem.*, **2**, 289–302.
 58. Franceschi,F. and Duffy,E.M. (2006) Structure-based drug design meets the ribosome. *Biochem. Pharmacol.*, **71**, 1016–1025.
 59. Maxwell,M.M. (2009) RNAi applications in therapy development for neurodegenerative disease. *Curr. Pharm. Des.*, **15**, 3977–3991.

60. Coyne,A.G., Scott,D.E. and Abell,C. (2010) Drugging challenging targets using fragment-based approaches. *Curr. Opin. Chem. Biol.*, **14**, 299–307.
61. Wu,W. (2010) MicroRNA: potential targets for the development of novel drugs? *Drugs R D*, **10**, 1–8.
62. Collie,G.W. and Parkinson,G.N. (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.*, **40**, 5867–5892.
63. Czech,M.P., Aouadi,M. and Tesz,G.J. (2011) RNAi-based therapeutic strategies for metabolic disease. *Nat. Rev. Endocrinol.*, **7**, 473–484.
64. Deigan,K.E. and Ferre-D'Amare,A.R. (2011) Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs. *Acc Chem. Res.*, **44**, 1329–1338.
65. Wilson,W.D. and Li,K. (2000) Targeting RNA with small molecules. *Curr. Med. Chem.*, **7**, 73–98.
66. Tor,Y. (2003) Targeting RNA with small molecules. *Chembiochem*, **4**, 998–1007.
67. Hermann,T. (2005) Drugs targeting the ribosome. *Curr. Opin. Struct. Biol.*, **15**, 355–366.
68. Wu,L., Pan,J., Thoroddsen,V., Wysong,D.R., Blackman,R.K., Bulawa,C.E., Gould,A.E., Ocain,T.D., Dick,L.R., Errada,P. et al. (2003) Novel small-molecule inhibitors of RNA polymerase III. *Eukaryot Cell*, **2**, 256–264.
69. Richter,S.N. and Palu,G. (2006) Inhibitors of HIV-1 Tat-mediated transactivation. *Curr. Med. Chem.*, **13**, 1305–1315.
70. Thomas,J.R. and Hergenrother,P.J. (2008) Targeting RNA with small molecules. *Chem. Rev.*, **108**, 1171–1224.
71. Giri,P. and Kumar,G.S. (2009) Molecular aspects of small molecules-poly(A) interaction: an approach to RNA based drug design. *Curr. Med. Chem.*, **16**, 965–987.
72. Aboul-ela,F. (2010) Strategies for the design of RNA-binding small molecules. *Future Med. Chem.*, **2**, 93–119.
73. Neelakandan,K., Babu,P. and Nair,S. (2012) Emerging roles for modulation of microRNA signatures in cancer chemoprevention. *Curr. Cancer Drug Targets*, **12**, 716–740.
74. Guan,L. and Disney,M.D. (2012) Recent advances in developing small molecules targeting RNA. *ACS Chem. Biol.*, **7**, 73–86.
75. Kumar,G.S. (2012) RNA targeting by small molecules: binding of protoberberine, benzophenanthridine and aristolochia alkaloids to various RNA structures. *J. Biosci.*, **37**, 539–552.
76. Liu,X., Liu,L., Xu,Q., Wu,P., Zuo,X. and Ji,A. (2012) MicroRNA as a novel drug target for cancer therapy. *Expert Opin. Biol. Ther.*, **12**, 573–580.
77. Mackowiak,G., Jens,S.D., Maaskola,J., Kuntzagk,A., Rajewsky,N., Landthaler,M. and Dieterich,C. (2012) doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.*, **40**, D180–D186.
78. Augustin,R., Endres,K., Reinhardt,S., Kuhn,P.H., Lichtenenthaler,S.F., Hansen,J., Wurst,W. and Trumbach,D. (2012) Computational identification and experimental validation of microRNAs binding to the Alzheimer-related gene ADAM10. *BMC Med. Genet.*, **13**, 35.
79. Cao,Y., Wu,J., Liu,Q., Zhao,Y., Ying,X., Cha,L., Wang,L. and Li,W. (2010) sRNATarBase: a comprehensive database of bacterial sRNA targets verified by experiments. *RNA*, **16**, 2051–2057.
80. Cerami,E.G., Gross,B.E., Demir,E., Rodchenkov,I., Babur,O., Anwar,N., Schultz,N., Bader,G.D. and Sander,C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.
81. Chen,X., Liu,M.X. and Yan,G.Y. (2012) RWRMADA: predicting novel human microRNA-disease associations. *Mol. Biosyst.*, **8**, 2792–2798.
82. Cho,S., Jang,I., Jun,Y., Yoon,S., Ko,M., Kwon,Y., Choi,I., Chang,H., Ryu,D., Lee,B. et al. (2013) MiRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Res.*, **41**, D252–D257.
83. Cannone,J.J., Subramanian,S., Schnare,M.N., Collett,J.R., D'Souza,L.M., Du,Y., Feng,B., Lin,N., Madabusi,L.V., Muller,K.M. et al. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**, 2.
84. Cvitkovic,I. and Jurica,M.S. (2013) Spliceosome database: a tool for tracking components of the spliceosome. *Nucleic Acids Res.*, **41**, D132–D141.
85. George,A.D. and Tenenbaum,S.A. (2011) Web-based tools for studying RNA structure and function. *Methods Mol. Biol.*, **703**, 67–86.
86. Pang,K.C., Stephen,S., Dinger,M.E., Engstrom,P.G., Lenhard,B. and Mattick,J.S. (2007) RNAdB 2.0—an expanded database of mammalian non-coding RNAs. *Nucleic Acids Res.*, **35**, D178–D182.
87. Supekar,K. (2010) 2D Chemical Structure Editor. *Masters Thesis*. University of Pune.
88. Wang,R., Fu,Y. and Lai,L. (1997) A new atom—additive method for calculating partition coefficients. *J. Chem. Inf. Comput. Sci.*, **37**, 615–621.
89. Cheng,T., Zhao,Y., Li,X., Lin,F., Xu,Y., Zhang,X., Li,Y., Wang,R. and Lai,L. (2007) Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model.*, **47**, 2140–2148.
90. Edwards,M.P. and Price,D.A. (2010) Role of physicochemical properties and ligand lipophilicity efficiency in addressing drug safety risks. *Annu. Rep. Med. Chem.*, **45**, 381–391.
91. Lipinski,C.A., Lombardo,F., Dominy,B.W. and Feeney,P.J. (2012) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliver Rev.*, **64**, 4–17.
92. Corey,R.B. and Pauling,L. (1953) Molecular models of amino acids, peptides, and proteins. *Rev. Sci. Instrum.*, **24**, 621–627.
93. Koltun,W.L. (1965) Space filling atomic units and connectors for molecular models, Washington, DC, USA, U. S. Patent 3170246.
94. Backman,T.W., Cao,Y. and Girke,T. (2011) ChemMine tools: an online service for analyzing and clustering small molecules. *Nucleic Acids Res.*, **39**, W486–W491.
95. Hawkins,P.C., Skillman,A.G., Warren,G.L., Ellingson,B.A. and Stahl,M.T. (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.*, **50**, 572–584.
96. O'Boyle,N.M., Banck,M., James,C.A., Morley,C., Vandermeersch,T. and Hutchison,G.R. (2011) Open Babel: an open chemical toolbox. *J. Cheminform.*, **3**, 33.
97. Lewis,K. (2013) Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.*, **12**, 371–387.
98. Lipinski,C.A., Lombardo,F., Dominy,B.W. and Feeney,P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **46**, 3–26.
99. Lagorce,D., Maupetit,J., Baell,J., Sperandio,O., Tuffery,P., Miteva,M.A., Galons,H. and Villoutreix,B.O. (2011) The FAF-Drugs2 server: a multistep engine to prepare electronic chemical compound collections. *Bioinformatics*, **27**, 2018–2020.
100. Baell,J.B. and Holloway,G.A. (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.*, **53**, 2719–2740.
101. Downs,G.M. and Barnard,J.M. (2002) Clustering methods and their uses in computational chemistry. In: Lipkowitz,B.K. and Boyd,B.D. (eds), *Reviews in Computational Chemistry*, Vol 18. John Wiley and Sons, Wiley-VCH, New Jersey, USA, pp. 1–40.
102. Wild,D.J. and Blankley,C.J. (1999) VisualiSAR: a web-based application for clustering, structure browsing, and structure-activity relationship study. *J. Mol. Graphics Model.*, **17**, 85–89.
103. Nicholls,A., McGaughey,G.B., Sheridan,R.P., Good,A.C., Warren,G., Mathieu,M., Muchmore,S.W., Brown,S.P., Grant,J.A., Haigh,J.A. et al. (2010) Molecular shape and medicinal chemistry: a perspective. *J. Med. Chem.*, **53**, 3862–3886.